

# EFFICIENT IMAGE RETRIEVAL USING REGION BASED IMAGE RETRIEVAL

Niket Amoda and Ramesh K Kulkarni

Department of Electronics and Telecommunication Engineering,  
Vivekanand Institute of Technology, University of Mumbai  
M.G. Road Fort, Mumbai, India  
niketamoda@gmail.com  
rk1\_2002@yahoo.com

## **ABSTRACT**

*Early image retrieval techniques were based on textual annotation of images. Manual annotation of images is a burdensome and expensive work for a huge image database. It is often introspective, context-sensitive and crude. Content based image retrieval, is implemented using the optical constituents of an image such as shape, colour, spatial layout, and texture to exhibit and index the image. The Region Based Image Retrieval (RBIR) system uses the Discrete Wavelet Transform (DWT) and a k-means clustering algorithm to segment an image into regions. Each region of the image is represented by a set of optical characteristics and the likeness between regions and is measured using a particular metric function on such characteristics.*

## **KEYWORDS**

*Content based image retrieval, K-Means Algorithm, Discrete Wavelet Transform, Region Based Image Retrieval.*

## **1. INTRODUCTION**

Early image retrieval techniques were based on textual annotation of images. By using text descriptions, images can be arranged by topical or syntactic classification to simplify navigation and browsing on the basis of standard Boolean queries. It was well admitted that a more accomplished and direct method to exhibit and index optical information would be based upon the fundamental characteristics of the images themselves.

Content based image retrieval, is implemented using the optical constituents of an image such as shape, colour, spatial layout, and texture to exhibit and index the image. In ideal content based image retrieval systems, the optical characteristics of the images in the database are extracted and illustrated by multi-dimensional feature vectors. The feature vectors of the images present in the database result in formation of a feature database.

For image retrieval, the users feeds example images or sketched figures to the retrieval system. The system then converts these examples images into its internal representation of feature vectors. The similarities / distances between the feature vectors of the query example or sketch and those of the images available in the image database are then computed and retrieval is performed using an indexing scheme. The indexing scheme gives an efficient approach to examine the image database.

Present-day retrieval systems have included users' relevance feedback to adjust the retrieval process in order to create perceptually and semantically more accurate retrieval results.

An optical content descriptor could be either local or global. A global descriptor uses the optical characteristics of the whole image, whereas a local descriptor uses the optical characteristics of regions or objects to describe the image content. In order to acquire the local visual descriptors, an image is often segmented into parts first.

Some of the widely used techniques for extracting color, texture, shape and spatial relationship features from images are now described briefly.

Instead of exact matching, content based image retrieval systems calculate the visual similarities between a query image and the images in a database. The result of the retrieval is not just a single image but, a list of images arranged according to their similarities with the query image. Different types of similarity/distance measures will influence the performances of an image retrieval system considerably. Commonly used similarity measures are: Mahalanobis Distance, Euclidean Distance and Bhattacharyya Distance.

One of the important issues in content-based image retrieval is effective indexing and faster image retrieval on the basis of optical characteristics. Since the feature vectors of images tend to have high dimensionality and hence they are not suitable for conventional indexing structures, dimension reduction is usually done before setting up an effective indexing scheme. Principal component analysis (PCA) is also one of the methods commonly used for dimension reduction. In this method we linearly map input data to a coordinate space, so that the axes are aligned to reflect the maximum variations in the data.

After reduction of dimension, the indexing of multi-dimensional data is done. There are various methods available for this purpose, such as R-tree (particularly, R\*-tree), K-d-B tree, Linear quad-trees and Grid files etc.

## **2. WAVELET TRANSFORMS**

The signals produced from natural sources such as digital images often have non-stationary attributes i.e. their content are variable in time or space.

Frequency analysis of stationary signals can be effectively achieved by projecting the signal onto a set of infinite spatial extent basis functions using the Fourier transform:

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-2j\pi ft} dt \quad (1)$$

where  $X(f)$  represents the global frequency of the signal. Similarly, effective frequency analysis of non-stationary signals can be achieved by projecting the signal onto a set of spatially localized basis functions using the wavelet transform.

$$\Psi(a, b) = \int_{-\infty}^{+\infty} x(t) \Psi_{ab}^*(t) dt \quad (2)$$

where  $a, b \in \mathbb{R}$  and  $\Psi_{ab}(t)$  is the translated and scaled version of the mother wavelet  $\psi(t)$  given by

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad (3)$$

Different choices of  $a$  and  $b$  result in the many different possible wavelet bases at different scales and translations.

#### A. Discrete Wavelet Transform (DWT)

The continuous convolution for discrete case the in the above equation is replaced by the following discrete summation:

$$c_{m,n} = \langle x, \Psi_{m,n} \rangle \equiv \sum_l x(l) \Psi_{m,n}(l) \quad (4)$$

where  $c_{m,n}$  are the wavelet coefficients. The implementation of the convolution of the scaling function with the signal is done at each scale through the iterative filtering of the signal with a low pass FIR filter  $h_n$ . At each scale the approximation coefficients  $a_{m,n}$  can be obtained using the following recursive relationship:

$$a_{m,n} = \sum_k h_{2n-k} a_{m-1,k} \quad (5)$$

where  $a_{0,n}$  is the sampled signal itself. In addition to this, if we use a related high pass FIR filter  $g_n$  the wavelet coefficients is obtained using the further recursive relation:

$$c_{m,n} = \sum_k g_{2n-k} a_{m-1,k} \quad (6)$$

In reconstruction of the original signal, the analysis filters can be selected from a biorthogonal set having a related set of synthesis filters. The synthesis filters  $\tilde{g}$  and  $\tilde{h}$  can be used to perfectly rebuild the signal using the reconstruction formula:

$$a_{m-1,l}(f) = \sum_n [\tilde{h}_{2n-l} a_{m,n}(f) + \tilde{g}_{2n-l} c_{m,n}(f)] \quad (7)$$

#### B. Extension of DWT to Two Dimensions

To extend the wavelet transform to two dimensions it is just necessary to separately filter and down sample in the horizontal and vertical directions. This produces four subbands at each scale. Denoting the horizontal frequency and followed by the vertical frequency, generates high-high (HH), high-low (HL), low-high (LH) and low-low (LL) image subbands. We can recursively apply the same scheme to the low-low subband a multiresolution decomposition can be obtained.

Figure 1(a) shows the normal layout of such a wavelet decomposition. The subbands are sensitive to frequencies at that scale and the LH, HL and HH subbands are sensitive to vertical, horizontal and diagonal frequencies respectively. Figure 1(b) shows a DWT decomposition of a texture image. This image shows the variation between wavelet subbands highlighting the scale and orientation selectivity of the transform. Also, Figure 1(d) shows a DWT decomposition of the Barbara test image (shown in Figure 1(c)). This image also shows the scale and orientation selectivity of the DWT. High energy subband regions pick out the texture content at different scales and orientations.

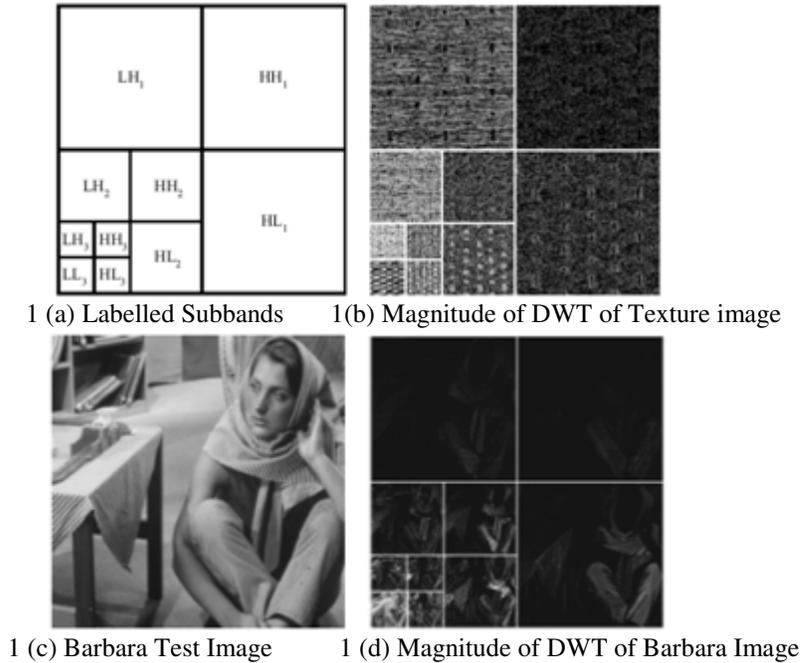


Fig: 1 Two Dimensional Wavelet Transform

### C. Extension of DWT to Two Dimensions

Haar function is one of the oldest and simplest example of a mother wavelet function. It is composed of a pair of rectangular pulses:

$$\Psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

A standard two-dimensional Haar wavelet decomposition of an image is very simple to code. It involves a one-dimensional decomposition on each row of the image, followed by a one-dimensional decomposition on each column of the result.

## 3. REGION BASED IMAGE RETRIEVAL

The Region Based Image Retrieval (RBIR) system uses the Discrete Wavelet Transform (DWT) and a k-means clustering algorithm to segment an image into regions. Each region is defined by

means of a set of characteristics and the likeness between the regions is measured using a specific metric function on such characteristics. The implementation of RBIR can be divided into two parts: Image Pre-processing and Image Retrieval.

### A. Image Pre-processing

The detailed algorithm for the pre-processing stage is given below.

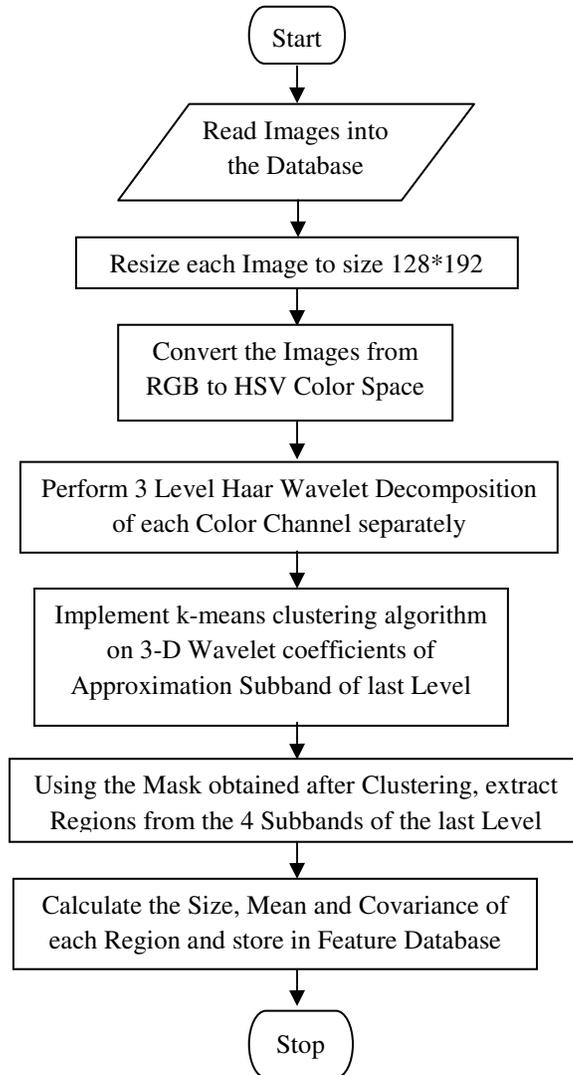


Fig: 2 Flow Diagram

The algorithm requires that all the images in the database and the query image be of the same size. A size of 128\*192 was chosen for all of the images. If the image had a different size, it was first resized to 128\*192 and then the pre-processing operations were carried out on the image.

In the RBIR application, each image is divided into the corresponding color channels (i.e. H, S and V) and the DWT was applied separately to each color channel. The j-th Wavelet coefficient of subband B (B ∈ {LL, LH, HL, HH}), where L stands for “low” and H for “high”) and DWT level l is a 3-D vector i.e.

$$\omega_j^{l;B} = (\omega_{0j}^{l;B}, \omega_{1j}^{l;B}, \omega_{2j}^{l;B}) \quad (9)$$

Where each component refers to a color channel c (c ∈ {0, 1, 2}). The energy of  $w_j^{l;B}$  on the c and d channels is then defined as:

$$e_{cdj}^{l;B} = \omega_{cj}^{l;B} \cdot \omega_{dj}^{l;B} \quad (10)$$

When c=d,  $e_{ccj}^{l;B}$  is called the channel energy of channel c, whereas when c≠d,  $e_{cdj}^{l;B}$  is termed the cross-correlation energy between channels c and d. The energy vector

$$e_j^{l;B} = (e_{00j}^{l;B}, e_{01j}^{l;B}, e_{02j}^{l;B}, e_{11j}^{l;B}, e_{12j}^{l;B}, e_{22j}^{l;B}) \quad (11)$$

captures both color and texture information through channel and cross-correlation energies, respectively. This is known to be one of the most robust methods for the representation of texture features.

*K-Means Clustering:* The k-means algorithm segments the observations in the given data into k mutually exclusive clusters, and returns a vector of indices denoting which of the k clusters it has assigned each observation.

Each cluster is defined on the basis of its member objects and its centroid. The centroid for individual cluster is the point to which the sum of distances from all objects in that cluster or partition is minimized.

$$\sqrt{(p_1 - q_1)x^2 + (p_2 - q_2)x^2 + \dots + (p_n - q_n)x^2} = \sum_{i=1}^n (p_i - q_i) \quad (12)$$

While using the Euclidean distance, each centroid is calculated as the mean of the points present in the cluster. For example, if the data set is 3 D and the cluster has two points: X = (x1, x2, x3) and Y = (y1, y2, y3), then the centroid Z becomes Z = (z1, z2, z3), where z1 = (x1 + y1)/2 and z2 = (x2 + y2)/2 and z3 = (x3 + y3)/2.

The k-means algorithm is a two-phase iterative algorithm which minimizes the sum of point-to-centroid distances, summed over all k segments:

- 1) In the first phase we use what the literature often describes as "batch" updates, where each iteration contains reassigning points to their closest segment centroid, all at once, followed by recalculation of segment centroids. This phase supplies a fast but potentially only approximate solution as a beginning point for the second phase.
- 2) In the second phase we use what the literature often describes as "on-line" updates, where points are independently reassigned if doing so will decrease the sum of distances, and segment centroids are recalculated after each reassignment. Each iteration during the second phase contains one pass though all the points.

**B. Image Retrieval**

The detailed algorithm for the Image Retrieval phase is given below.

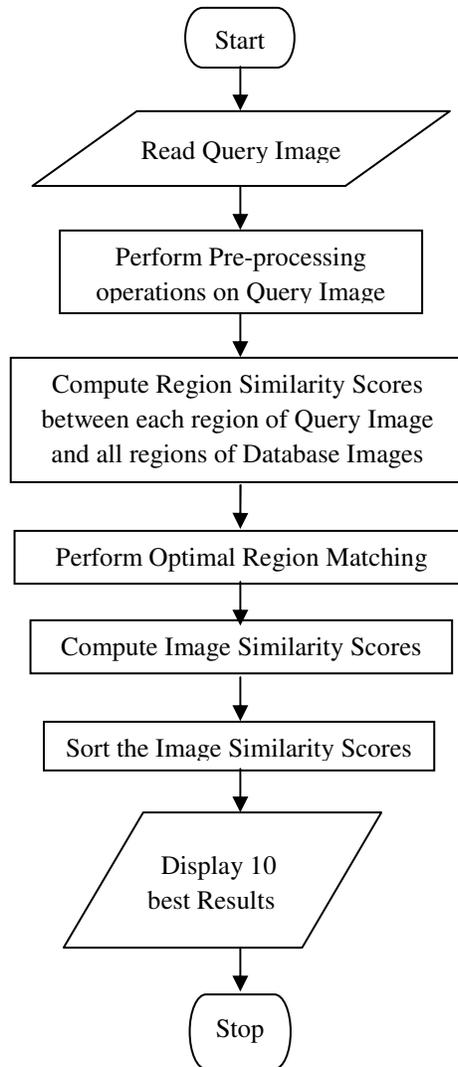


Fig: 3 Flow Diagram

After reading the query image, the same pre-processing operations of image resizing, RGB to HSV conversion, DWT decomposition, k-means clustering and feature extraction must be performed on the query image. At the end of the pre-processing operation, the sizes, mean vectors and covariance matrices of the regions of the query image would be obtained.

*Region Similarity* : The similarity between two regions  $R_{q,i}$  (represented by the feature vector  $[\mu_{R_{q,i}}, CR_{q,i,3}, size(R_{q,i})]$ ) of a query image  $I_q$  and  $R_{s,j}$  (represented by the feature vector  $[\mu_{R_{s,j}}, CR_{s,j,3}, size(R_{s,j})]$ ) of a database image  $I_s$  is computed as

$$r_{sim}(R_{q,i}, R_{s,j}) = d(R_{q,i}, R_{s,j}) \quad (13)$$

where  $d()$  is a distance function. The distance  $d(R_{q,i}, R_{s,j})$  between the regions  $R_{q,i}$  and  $R_{s,j}$  is a weighted sum, taken over the four frequency subbands, of the distances between color-texture descriptors, plus an additional term that takes into account the difference between the relative sizes of the two regions.

In the present work, all the frequency coefficients are equally weighed i.e.  $\gamma_B = 1$  for  $B \in \{LL, LH, HL, HH\}$ . The second term takes into account the difference in size between the regions by multiplying it with a coefficient that favors matches between large regions. The distance  $d_B(R_{q,i}, R_{s,j})$  between two regions on the frequency subband  $B$  is computed by using the Bhattacharyya Metric:

$$d_B(R_{q,i}, R_{s,j})^2 = \frac{1}{2} \ln \left( \frac{\left| \frac{C_{R_{q,i}}^{3;B} + C_{R_{s,j}}^{3;B}}{2} \right|}{\left| C_{R_{q,i}}^{3;B} \right|^{\frac{1}{2}} * \left| C_{R_{s,j}}^{3;B} \right|^{\frac{1}{2}}} \right) + \frac{1}{8} \left[ (\mu_{R_{q,i}}^B - \mu_{R_{s,j}}^B)^T * \left( \frac{C_{R_{q,i}}^{3;B} + C_{R_{s,j}}^{3;B}}{2} \right)^{-1} * (\mu_{R_{q,i}}^B - \mu_{R_{s,j}}^B) \right] \quad (14)$$

#### 4. IMPLEMENTATION OF RBIR

After the optimal region assignment has been performed, the next step is to compute the Image Similarity Score. This score is obtained by simply adding the region similarity scores of the matched regions. The final step is to sort the image similarity scores so obtained and then display the images having the least distance from the query image.

*Adding Images to the Database:* The steps involved in adding images to the database are:

- 1) Run the ‘Feature Database Generation Population’ program. A Matlab GUI appears as shown in Figure 4
- 2) Select the images from the folder Image\_Database and click on ‘Add’ to generate a feature database.
- 3) The images selected will appear in the listbox.
- 4) Click on the ‘Done’ after selecting the images to be added to the database.
- 5) Feature Database of the image will be generated.

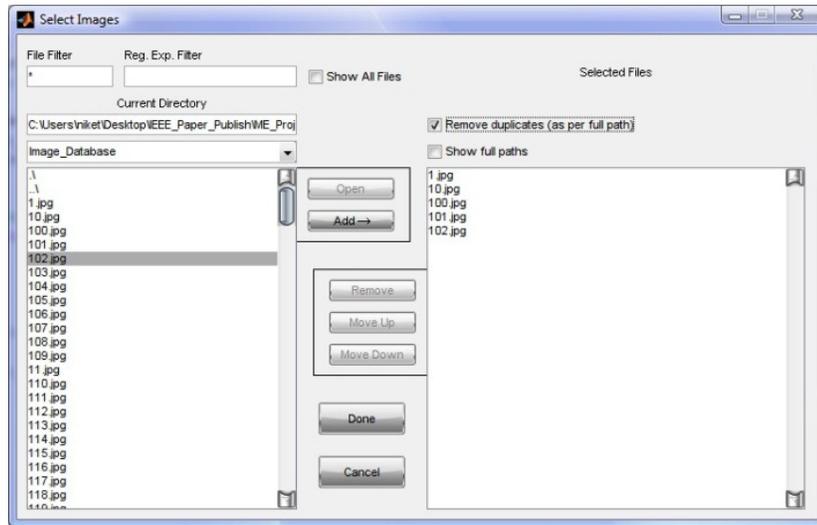


Fig: 4 Feature Database population

*Image Retrieval:* The steps involved in searching for images are:

- 1) Start the 'Region Based Image Retrieval' program. A GUI appears as shown in Figure 5 (a).
- 2) Select a query image from the folder Image\_Query.
- 3) The query image will be displayed Figure 5 (b) along with the 16 matches of the most similar images available in the database.

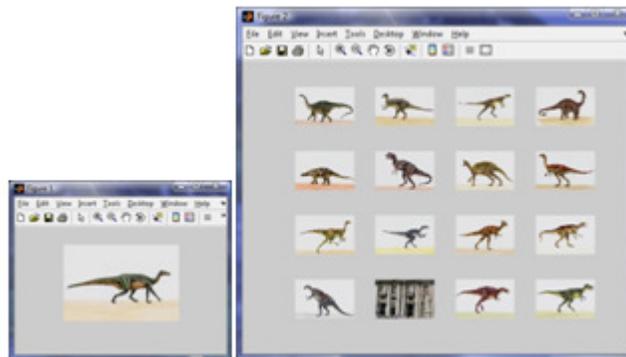


Fig: 5 (a)

Fig: 5 (b)

Fig: 5 Region Based Image Retrieval

## 5. RESULTS

To test the RBIR application, a database consisting of 180 general images, was used. The 180 images could roughly be categorized into 9 groups, each group consisting of 20 similar images. In addition 9 query images, each query corresponding to one of the groups were taken.

Figure 6 shows an example of the results achieved using the RBIR application. From a semantic point of view, the results obtained are particularly good i.e. all the images in this particular example are of horses.

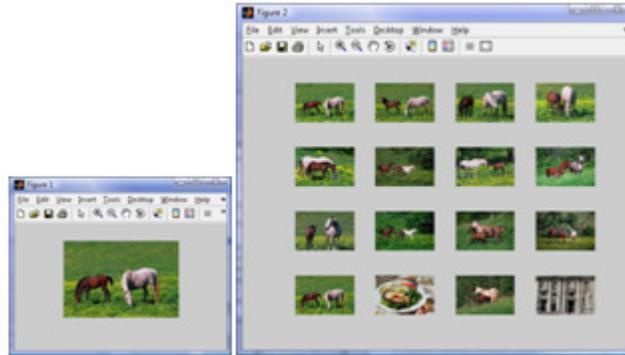


Fig: 6 RBIR Results

#### A. *Partial Match Queries*

A partial match query is a query that specifies only part of the image. In Figure 7, the query image is obtained by cropping a database image. As the results show, the RBIR application gave the complete image as the very first match.

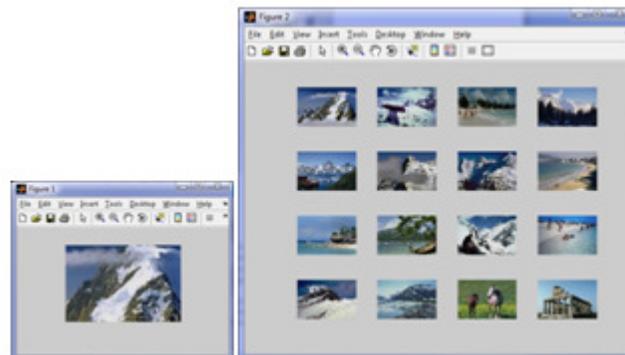


Fig: 7 Partial Match Querying

#### B. *Scanned Queries*

During the scanning process of query image, it may suffer artifacts such as poor resolution, misregistration, color shift and dithering effects.. To consider the effect of scanned images on the retrieval effectiveness, the query image was first printed and then subsequently scanned. The scanned image appeared fuzzier, darker and slightly misregistered compared to the original. Figures 8 (a) and (b) displays the results obtained by the original query image and the scanned query image respectively. It can be observed that there is a slight degradation in the quality of the results obtained.

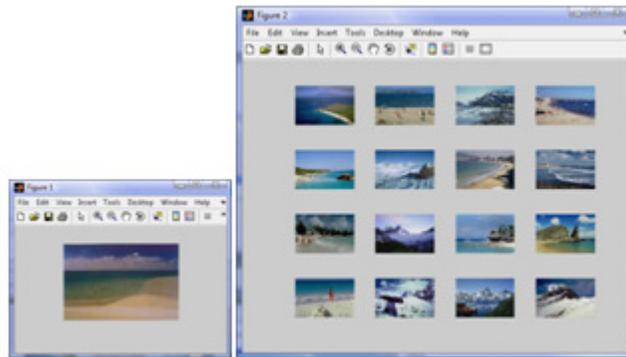


Fig: 8 (a) Result with Original Query

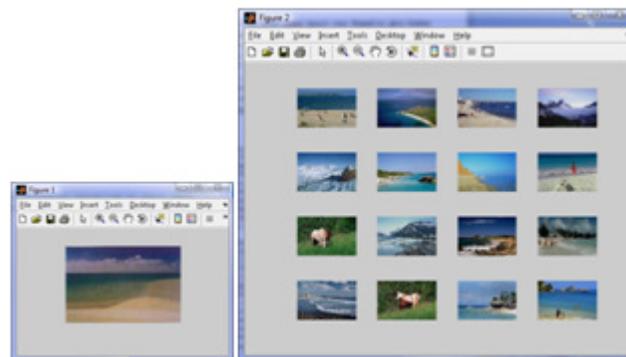


Fig: 8 (b) Result with Scanned Query

Fig: 8 Scanned Queries

### C. Difficult Queries

The effectiveness of the RBIR application is confirmed when considering “difficult” queries, i.e. queries having a low number of similar images in the database. Figure 9 shows the results for a query having only two similar images in the database. The RBIR system is able to retrieve both of these images.

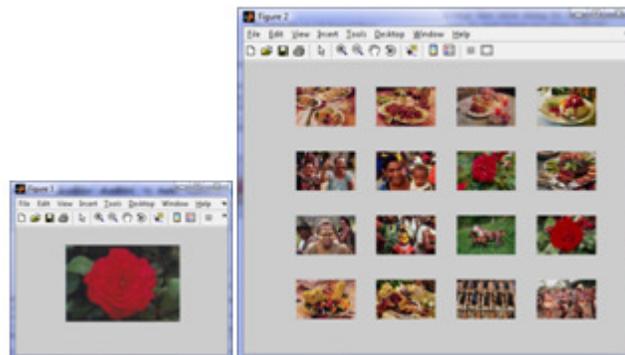


Fig: 9 Difficult Queries

#### **D. Search Time**

For an image retrieval application, the time taken for retrieval is an extremely important parameter. On the other hand, the time taken for pre-processing is not as important since the pre-processing operations have to be carried out only once.

The first entry in table below shows the average time taken to perform the pre-processing operations on the database of 180 images of size 128\*192. The second entry shows the average time taken during the image retrieval phase. Here again, the query image was of size 128\*192.

The experimental setup consisted of a computer with a 2.4 GHz Intel(R) Core(TM)2 Duo CPU and 3 GB of DDR2 PC RAM running MATLAB 7 on Windows 7 Eternity.

Pre-processing Stage	25.2 seconds
Image Retrieval Stage	1.7 seconds

## **6. CONCLUSIONS**

Although the HSV color space was found to give better results compared to the RGB color space in, in our experiments the RGB and HSV color spaces were found to give almost equivalent results. Eventually, it was decided to use the HSV color space because it gave better results than the RGB color space in case of “difficult queries”. The figure below shows that when using the RGB color space, only one of the two matches was retrieved. On the other hand, in the HSV color space, both the matches were retrieved.

## **REFERENCES**

- [1] D.Lowe, “Object recognition from local scale-invariant features,” in ICCV, 1999, pp. 1150–1157.
- [2] Y.J.Zhang “A survey on evaluation methods for image segmentation”, Pattern Recognition 29 (8) (1996) 1335 - 1340
- [3] A.Jain, “Data clustering: 50 years beyond k-means,” Pattern Recognition Letters, vol. 31, no. 8, pp. 651 – 666, June 2010.
- [4] W.Zhao, H.Ma, Q.He, "Parallel K-Means Clustering Based on MapReduce," in: Cloud Computing, vol. 5931, pp. 674-679, 2009.
- [5] W.D.Arthur, S. Vassilvitskii, “K-means++: the Advantages of careful seeding,” in Proc. 2007 Symposium on Discrete Algorithms, pp.1027-1035.
- [6] Rafael C. Gonzalez, Richard E. Woods, " Digital Image Processing" , Second Edition, Prentice Hall Upper Saddle River, New Jersey 07458, TA1632.G66 2001, 698-740
- [7] Fast Multiresolution Image Querying, International Conference on Computer Graphics and Interactive Techniques, 1995: Charles E.Jacobs, Adam Finkelstein, David H. Salesin

- [8] Content-based Image Retrieval, A report to the JISC Technology Applications Programme, 1999: John Eakins, Margaret Graham
- [9] Fundamentals of Content-based Image Retrieval, Multimedia Information Retrieval and Management - Technological Fundamentals and Applications, Springer, 2002: Dr.Fuhui Long, Dr. Hongjiang Zhang, Prof. David Dagan Feng
- [10] Image Retrieval – Current techniques, Promising directions and Open issues, Journal of Visual Communication and Image Representation, 1999: Yong Rui, Thomas S.Huang, Shih-Fu Chang
- [11] Wavelet Based Texture Analysis and Segmentation for Image Retrieval and Fusion, Thesis, University of Bristol, 2002: Paul R. Hill
- [12] WINDSURF: A Region Based Image Retrieval System, Proceedings of the 10th International Workshop on Database & Expert Systems Applications, 2000: Ilaria Bartolini, Paolo Ciaccia, Marco Patella
- [13] P.Felzenszwalb, R.Girshick, D.McAllester, and D.Ramanan, “Object detection with discriminatively trained part based models,”in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, 2010.

## Authors

Niket Amoda received his B.E. in Electronics & Communication Engineering from Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal, India, in 2006. He completed his Diploma in Advance Computing from ACTS, Pune, India in 2008 and is M.E. scholar in Department of Electronics & Telecommunication Engineering, Vivekanand Education Society's Institute of Technology (VESIT), University of Mumbai, India. He worked as a Software Development Engineer in Ford Motor Company, India from 2008 to 2009. He is currently working as a Assistant Professor in Electronics & Telecommunication Engineering Department of KC College of Engineering & Management & Research, Thane, India, since 2010 till date. He has many publications in international journals and international conferences. His research area is Image processing. He is also engaged in SAP certification from SAP Germany.



Dr. R. K. Kulkarni completed his PHD from, National Institute of Technology, Rourkela, Orissa, India. He received his bachelor degree in Electronics & Communication from Mysore University and master degree in Digital Electronics form Karnataka University, Karnataka. He has many publications in international journals and international conferences. His research area is Image processing, Non – Linear filters, and Digital signal processing.

